

学校编码: 10384

分类号_____密级_____

学号: 21720081152655

UDC _____

厦 门 大 学

硕 士 学 位 论 文

药物不良反应层级系统的构建与分析

Construction and Statistics of ADR Classification System

(ADReCS)

潘 文

指导教师姓名: 纪志梁 教授

专 业 名 称: 生物化学与分子生物学

论文提交日期: 2011 年 05 月

论文答辩时间: 2011 年 06 月

学位授予日期: 2011 年 月

2011 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

目录

中文摘要	I
英文摘要	II
第一章 前言	1
1.1 药物不良反应	1
1.1.1 药物不良反应定义	1
1.1.2 药物不良反应分类	1
1.1.3 药物不良反应研究意义	2
1.2 药物不良反应术语库的现状	4
1.2.1 药物不良反应术语库简介	4
1.2.2 WHO-ART 与 MedDRA 术语库的构成	5
1.3 本论文的思路, 目的和意义	8
第二章 材料和方法	10
2.1 研究流程	10
2.2 数据信息的采集	11
2.2.1 药物 ADR 信息获取	11
2.2.2 药物其他信息的获取	12
2.3 ADR 信息处理	13
2.3.1 ADR 词条的标准化处理	13
2.3.2 ADR 描述语分级系统的建立	15
2.3.2.1 层级结构关系	17
2.3.2.2 层级结构 ID 赋予	18
2.3.3 ADR 同义词的引入	20
2.4 ADR 分级系统术语库 ADReCS 的建成	23
2.5 ADR 药物与相关 ADR 的对应	26
2.6 药物不良反应分级系统数据库的构建	26
2.6.1 需求分析	26
2.6.2 数据库框架设计	27
2.6.3 数据库配置	30
2.6.4 数据库各模块的实现	31
2.6.4.1 查询模块的实现	32
2.6.4.2 浏览模块的实现	36
2.6.4.3 记录提交模块的实现	37
2.6.5 数据库构建小结	38
第三章 结果和讨论	39
3.1 ADReCS 数据库数据记录	39
3.2 ADReCS 数据库数据分析	39

3.2.1 ADReCS 层级结构数据分析	39
3.2.1.1 ADReCS 四级词的分类概况.....	39
3.2.1.2 ADReCS 四级词的 ID 统计.....	41
3.2.2 药物-ADR 关系分析	46
3.2.2.1 ADReCS 中药物诱导的 ADR 统计	46
3.2.2.2 ADReCS 中 ADR 相关的药物统计	49
第四章 结论和展望	53
参考文献	56
致谢	60

Table of Contents

Abstract in Chinese	I
Abstract in English	II
Chapter 1 Introduction	1
1.1 Adverse Drug Reaction	1
1.1.1 ADR definition	1
1.1.2 ADR classification	1
1.1.3 Significance of ADR research	2
1.2 ADR terminology	4
1.2.1 About ADR terminology	4
1.2.2 WHO-ART and MedDRA	5
1.3 Idea, objective and significance of this article	8
Chapter 2 Materials and methods	10
2.1 Research process	10
2.2 Data sources	11
2.2.1 Data source of ADR information	11
2.2.2 Data source of other information	12
2.3 Data manipulation	13
2.3.1 Unification of ADR key words	13
2.3.2 Classification of ADR terms	15
2.3.2.1 Relationship between different levels	17
2.3.2.2 ID assignment of ADR terms	18
2.3.3 ADR synonyms extraction	20
2.4 Construction of the ADR Classification System	23
2.5 Mapping ADR and related drugs	26
2.6 Construction of ADReCS database	26
2.6.1 Requirements of database	26
2.6.2 Design of database	27
2.6.3 Configuration of database	30
2.6.4 Building modules in ADReCS database	31
2.6.4.1 Construct the searching module	32
2.6.4.2 Construct the browsing module	36
2.6.4.3 Construct the uploading module	37
2.6.5 Summary of Construction of ADReCS database	38
Chapter 3 Results and Discussion	39
3.1 Record information of ADReCS database	39
3.2 Statistics and analysis of ADReCS database	39

3.2.1	Statistics of ADReCS database.....	39
3.2.1 .1	Classification of ADReCS	39
3.2.1 .2	statistics of IDs of 4 th level terms	41
3.2.2	analysis of drug-ADR pairs.....	46
3.2.2.1	Numbers of ADRs per drug	46
3.2.2.2	Numbers of drugs per ADR	49
Chapter 4	Conclusion and prospect.....	53
Reference	56
Acknowledgement	60

摘 要

药物在发挥药效的同时,往往会产生药物不良反应(Adverse Drug Reaction, ADR)。基于其发生的普遍性和危害的严重性, ADR 已成为困扰当今世界医疗卫生领域的一个重要问题,也是目前制约新药开发和上市后临床应用的一大难题。近年来,计算机辅助毒理学的发展为 ADR 的机理研究和预测提供了新的方法。然而,由于缺乏较好的 ADR 标准化及分级系统,计算机辅助 ADR 研究受到了严重制约。

本论文通过采集 DailyMed 网站的药物 ADR 信息,参照 ICH 国际医学用语词典(MedDRA)将 ADR 词条标准化并构建了 ADR 分级系统术语库(ADR Classification System, ADReCS)的 1.0 版本。该系统也整合了一系列药物信息,并建立药物-ADR 关系,最终以网络数据库的形式展示(目前,尚未公开发布)。

目前为止,Andres 术语库包含一级词 26 个,二级词 327 个,三级词 1,213 个,四级词 5,398 个。ADReCS 数据库包含 1,159 个单一有效成分药物 ADR 信息,涉及 96,779 个药物-ADR 关系。对 ADReCS 中药物与 ADR 信息分析结果显示:72% 的药物诱导 10-100 种 ADR,诱导 ADR 最多的药物为神经系统相关药物,皮肤科用药诱导的 ADR 普遍较少。77% 的 ADR 仅在 10 种以下的药物发生,在药物中发生频率最高的 3 种 ADR 是反胃,头痛和呕吐。胃肠道系统,神经系统,皮肤及皮下组织系统,血管系统这 4 个器官系统是药物不良反应最常发生的系统。

药物不良反应层级系统 ADReCS 的建立将会为 ADR 机理研究及预测提供有力的数据支持,进一步促进计算机辅助毒理学和药物设计的发展。

关键词: 药物不良反应 ; 分级系统; 数据库

ABSTRACT

Adverse drug reaction (ADR) is a significant problem for new drug development and chemotherapy. Traditional ADR assessment methods cost huge capital and time. At the meanwhile, their performance is unsatisfied. The recent development of computer-aided toxicity research provides an alternative way for ADR mechanism research and ADR assessment. However, the shortness of proper ADR standard and classification system limits applications of computational toxicity.

In this thesis, we collected ADR information mainly through *DailyMed* Database and took the *MedDRA* as the reference to standardize all ADR entries to construct an ADR classification system (ADReCS). Moreover, the ADReCS provides connection between drug and ADR. The ADReCS was presented in a form of database and a series of interactive user interfaces were also designed. (At present, ADReCS has not been published yet.)

The current version of ADReCS database contained 26 1st level ADR terms, 327 2nd level ADR terms, 1,213 3rd level ADR terms and 5,398 4th level ADR terms. It was estimated that the ADReCS database contained 96,779 drug-ADR pairs and information of 1,159 single active ingredient drugs. Furthermore, 72% of drugs have about 10 to 100 different ADRs. Drugs of nervous system class have most ADRs while drugs of dermatological class usually have few ADRs. 77% of ADRs occur in less than 10 drugs. Nausea, headache and vomiting are most often ADRs that occur for most drugs. ADRs of gastrointestinal disorders, nervous system disorders, skin and subcutaneous tissue disorders and vascular disorders happen mostly frequently.

In summary, ADReCS will be helpful for future computational toxicity and drug development.

Keywords: ADR; classification system; database

第一章 前言

药物指能影响机体生理、生化和病理过程，用以预防、诊断、治疗疾病的化学物质。所有药物进入机体内，经过 ADME 过程，对人体产生各种药理作用，在发挥药效的同时，往往也会产生有害反应，通常称之为药物不良反应（Adverse Drug Reactions, ADR）。这些药物不良反应会影响药物的治疗效果，甚至危及患者生命。

1.1 药物不良反应

1.1.1 药物不良反应定义

根据 WHO 国际药物监测合作中心的定义，药物不良反应（Adverse Drug Reactions, ADR）是指正常剂量的药物用于预防、诊断、治疗疾病或调节生理机能时出现的非预期有害反应^[1]。该定义排除有意的或意外的过量用药及用药不当引起的反应。换言之，药品不良反应可能是药物作用机制非期待性的直接延伸。

1.1.2 药物不良反应分类

药物不良反应（ADR）发生的机理复杂，包括副作用、毒性反应、过敏反应、特异质反应和致畸作用等，这使得其分类方法多种多样。目前通常用的是 1977 年 Rawlins 和 Thompson 提出的 ADR 分类方法，按照 ADR 对药物有无剂量依赖性分为 A、B 两类反应^[2]。

A 型药物不良反应（Type A ADR）又称为剂量相关的不良反应（dose-related ADR），占药物不良反应的 70%-80%^[3]。该反应是由于药物的药理作用增强所致，其特点是可以预测的，通常与用药剂量相关，停药或减量后症状很快减轻或消失，发生率高，但死亡率低，其中副作用、毒性反应、过度效应等均属 A 型不良反应，例如：阿托品引起的口干，抗凝血药所致的出血等。A 型 ADR 发生原因主要有：药物代谢动力学方面原因，包括药物的吸收，分布，与血浆蛋白或组织结

合，生物转化及肾脏排泄；机体离子平衡受影响及受体调节异常。

B 型药物不良反应 (Type B ADR)，又称剂量不相关的不良反应 (non-dose-related ADR)，占药物不良反应事件总数的 20%~25%。它是一种与正常药理作用无关的异常反应，一般和剂量无关联，难于预测，发生率低，而死亡率高，常规毒理学筛选不能发现，其中药物变态反应和异质反应属 B 型不良反应，如氟烷引起的恶性高热，青霉素的过敏反应。B 型 ADR 发生原因主要有：药物因素，ADR 是由药物生产过程中使用的各种添加剂，如稳定剂、着色剂、赋形剂等，或化学合成中产生的杂质以及药品在贮存、运输过程中产生的氧化、分解、降解、聚合等产物所致，而非药物有效成分本身；患者自身机体因素，机体可能存在某些不易发现的异常和缺陷，一旦暴露在药物下，则会出现不良反应。

1.1.3 药物不良反应研究意义

长期以来，大多数常用药物对于患同样疾病的患者来说，有效率只占 30%-60%，不仅如此，对于部分患者来说还可能有严重的不良反应。20 世纪的典型重大药害事件，有 30 年代的磺胺酞剂事件^[4]和 60 年代的沙利度胺(反应停)事件^[5]。还有异丙肾上腺素喷剂引起致命性心律失常；氨基比林引起粒细胞减少；氯碘羟喹引起急性脊髓视神经病变、己烯雌酚引起下一代少女阴道癌；氯霉素和保泰松引起再生障碍性贫血等药物严重不良反应^[6]。对患者而言，药物不良反应严重影响了其健康和生活质量，给家庭和社会带来沉重的经济负担。据有关资料评估，中国每年约有 5 000 万人住院，其中至少 250 万人(20%)是因 ADR 住院，50 万人是严重的 ADR，每年死亡约 19 万人，从而增加医药费近 40 亿元。

严重的药物不良反应还使某些本来具有良好治疗前景的药物离开市场。1980 - 2001 年，美国 FDA 宣布，从美国市场上撤回了 20 种处方药，如舒洛芬、芬氟拉明、特非那定、苯丙醇胺、西立伐他汀等。据不完全统计，在 1960 - 1999 年，全球已有 121 个处方药因安全性问题，撤离市场。近年影响较大的撤市产品，2004 年 9 月，在全球范围内撤市治疗类风湿关节炎的 COX - 2 抑制剂罗非昔布(商品名：万络)；相继还有 COX - 2 抑制剂伐地考昔(Valdecoxib, Bextra) 和糖尿病治疗药物曲格列酮。2006 年 5 月，美国和加拿大撤市了喹诺酮类(沙星类)抗菌药品加替沙星片剂和注射剂；2007 年，胃肠道疾病治疗药物泽马可(马来酸替加色罗)撤市；治疗晚期肺癌药物易瑞沙(Iressa)在欧洲自动撤市；2008 年 10 月，上市仅 2 年的新型控

制体质量药物“利莫纳班”(选择性 CB1 大麻受体拮抗剂),由于神经系统的安
全性问题,在欧盟撤市,并终止了在我国进口注册申请和全球所有正在进行的国
际多中心临床试验。

如图 1.1 所示,开发一个新药是一项长期且复杂的工作,在其上市前就需经
历大量的临床试验检测其安全性。据调查,在美国上市一种新药平均要花费 5 亿
多美元,耗时约 12 ~ 15 年时间。研制的新药中只有不到 5 %能够进入临床前
研究阶段,然后又只有 2 %能进入临床试验阶段,开展 I 期临床试验的所有研
制药物有 80 %会在上市前淘汰出局。在新药审批之前的研发过程中,新药的通过
率在 1 / 5000 左右。这其中 1/3 的研发药物失败都是药物 ADR 导致的^[7]。
因而对于药物不良反应的预测将会明显缩短药品临床试验的时间周期,降低成
本,在整个研究过程中尽早淘汰不适合的化合物,大大加快新药研发的进程。

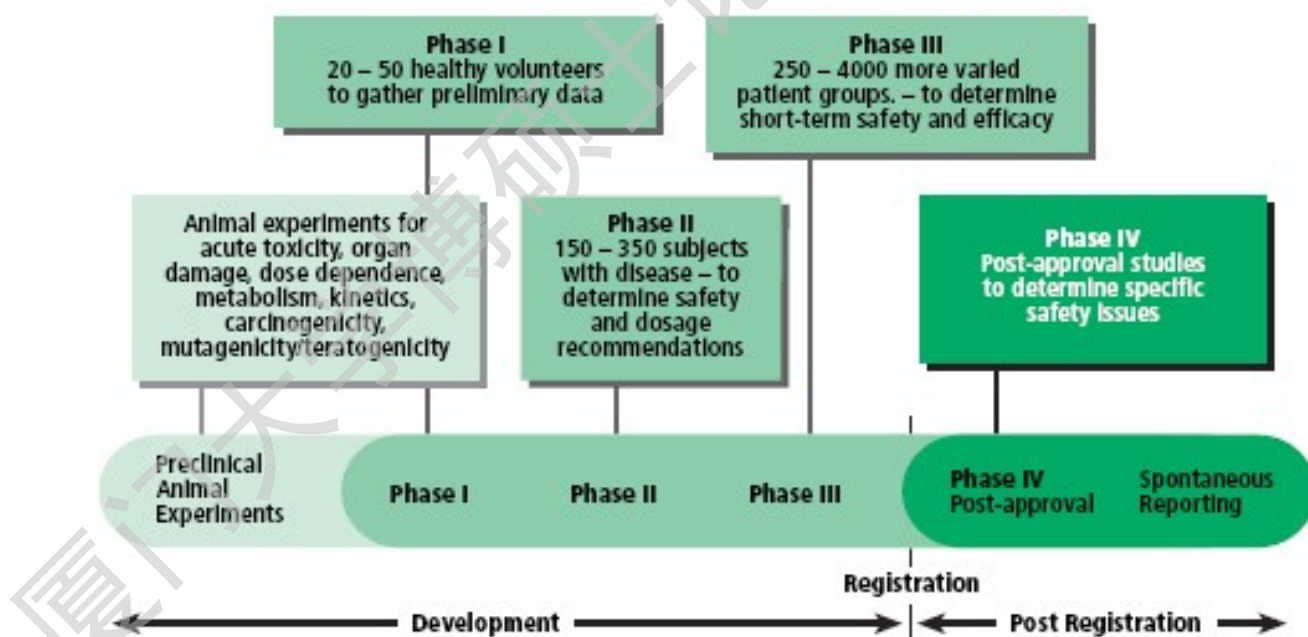


图 1.1 药物临床试验流程图

Fig.1.1 Clinical development of medicines

资料来源: World Health Organization, WHO Policy Perspectives on Medicines — Pharmacovigilance:

ensuring the safe use of medicines, October 2004

1.2 药物不良反应术语库的现状

1.2.1 药物不良反应术语库简介

鉴于药物不良反应的普遍性和严重性后果,世界各国相关组织都建立了ADR 监测报告体系,及时发现新的、严重的药品不良反应,以便国家药品监督管理部门及时对有关药品加强管理,避免同样药品、同样不良反应的重复发生,保护更多人的用药安全和身体健康。

药品不良反应自愿报告制度是目前药品上市后最广泛采用的监测手段。自愿报告制度是指由医疗卫生专业人员或消费者非医疗卫生专业人员将某种药品所引起的不良反应主动直接地呈报给药监机构、制药企业,或通过医药文献杂志等途径进行报道。其监测范围广,参与人员多,不受时间、空间的限制,是报告的主要信息源。自愿报告形式的多样性和来源于非医疗卫生专业人员的比例高,导致报告信息不完整和医学术语使用不规范^[8]。

为规范 ADR 事件报告用词以有效获取药物 ADR 事件信息,各国纷纷制定了各自的药物不良反应术语库,以利于相关数据库的建立和信息的检索。如美国使用药物与食品监督管理局(Food and Drug Administration, FDA)编订的副反应术语编码词典(The Coding Symbols for a Thesaurus of Adverse Reaction Terms, COSTART)及国际疾病分类词库(International Classification of Diseases, ICD)^[9];欧洲使用国际疾病分类词库(International Classification of Diseases, ICD)与世界卫生组织副反应术语集(World Health Organization Adverse Reaction Terminology, WHO-ART),分别用于报告药物上市前临床实验阶段出现的副作用,和上市后通过医院监测和信息呈报系统发现的副作用;日本使用日本副反应术语集(Japanese Adverse Reaction Terminology, J-ART)结合医学信息系统(Medical Information System (Japan), MEDIS)来呈报药物安全方面的信息。

不同国家采用的药物不良反应术语库不同,造成不同地区在药物上市前后的 ADR 事件呈报系统中纪录的数据形式不一,格式各异,不便于数据本身的检索分析、交叉引用。采用不同术语集的数据库在交换数据信息时,必须先转换成统一的术语集,这将耗费大量人力物力,并可能造成时间延误、信息丢失,给各方研究人员带来极大不便。为在世界范围内建立一个统一的用于药事管理为目的

的医学术语集,囊括与替代原有的医药产品开发用的各种词典(包括各类 ADR 词库),人用药物注册技术要求国际协调会议(International Conference on Harmonisation, ICH)于 1994 年正式提出建立一个国际医学用语词典(Medical Dictionary for Regulatory Activities, MedDRA),并指定其维护与客户服务机构(Maintenance and Support Services Organization, MSSO),对 MedDRA 进行维护、发行与进一步开发。2010 年 9 月 MSSO 发行的 MedDRA 13.1 包括了术语库 J-ART、COSART、HARTS(Hoechst Adverse Reaction Terminology System)以及 WHO-ART 等术语库中的几乎所有词汇。

目前,国际上 ADR 事件记录最为常用的术语库是 WHO-ART 和 MedDRA。由 UMC 维护的隶属于 WHO 的 Vigibase 数据库,就对这两个术语库兼收并蓄,并行使用。WHO-ART 作为 30 多年来 WHO 推荐使用的术语库,短小精悍、灵活实用。而新晋的 MedDRA 以其适用于药事管理各方面的详细的分类、丰富的信息、系统的层级而崭露头角,已经成为药品不良反应报告和制药工业的国际性的标准术语集,渐有取代 WHO-ART 之势。

1.2.2 WHO-ART 与 MedDRA 术语库的构成

WHO-ART 术语库由系统器官分类(System Organ Class, SOC)、高位语(High Level Term, HLT)、首选语(Preferred Term, PT)和包含语(Included Term)4 个描述药品不良反应的层级组成。包含了 32 个术语的系统器官分类是最高层级,其中 27 个 SOC 术语按照组织学原理进行分类,另外 5 个术语则分别为“全身性失调”、“耐受机制紊乱”、“用药部位副反应”、“次级术语”以及“中毒特异性术语”。高位语将相关或相近的首选语归属为同一类,以方便数据检索,但并非每个首选语都必然存在一个对应的高位语。同时,WHO-ART 中的高位语词也是其下级首选语词当中的一个,现有 180 个 HLT 术语。首选语,指各国上报 ADR 事件的推荐语言,即上报关键词,用于描述各种性质不同的副作用,目前已包含 PT 术语 2,085 个。包含语则是各个首选语的同义词形式,以适应不同地区不同机构的不同用词习惯,WHO-ART 中现有 3,445 个包含语。

如下表 1.4 所示,WHO-ART 以“记录号(4 位)+序列号(3 位)”的方式编码首选语与包含语,从属于同一个首选语的所有包含语与该首选语拥有同样的记录号,区别仅在于首选语的序列号为 001,而包含语则是 001 以外的序数递增,

从而保障每个词的编码的唯一性。WHO-ART 中高位语和系统器官分类语分别用 4 位数字编码,其中高位语编号选用其下属所有首选语中记录号最小者(如表 1.4 所示)。此种编码方式优点是各层级可扩展性强,从编码可看出内含语和首选语的从属关系,可分辨词条属于包含语/首选语或是高位语/系统器官分类语。

表 1.4 WHO-ART 结构及层级关系
Table 1.4 The structure of WHO-ART and relationship between different types of terms

ARECNO +SEQ	Included Terms	ARECNO +SEQ	Preferred Terms	HL Link	High Level Terms
0363 005	Acidos metabolic	0363 001	Acidosis	0363	Acidosis
0363 004	Bicarbonate reserve decrease				
0364 003	Lactate blood increase	0364 001	Acidosis lactic	0363	
0393 003	Ketoacidosis	0393 001	Ketosis	0363	
0393 004	Acetonuria				
0393 005	Aceton breath				
0393 006	Acetonaemia				

MedDRA 的分级系统更为细化,由 5 个层级构成,分别为:系统器官分类(System Organ Class, SOC)、高位组语(High Level Group Terms, HLGT)、高位语(High Level Terms, HLT)、首选语(Preferred Terms, PT)及低位语(Lowest Level Term, LLT)。MedDRA 的基本单位是首选语,结构特点与 WHO-ART 类似,与之词库相比,仅多了一个高位组语 HLGT 层级(低位语相当于 WHO-ART 中的包含语)。由于 MedDRA 是一个涉及到整个药事管理过程的标准术语集,故其 SOC 不仅包括了药物不良反应相关词,还包括“社会史”、“外科手术和医学治疗过程”等其他药事相关词。MedDRA 的词汇量远高于 WHO-ART,其首选语及低位语数量分别为 18,919 和 68,661 (MedDRA Version 13.1) 而 WHO ART 中的首选语和内含语数量仅分别为 2,085 和 3,445 (截止至 2005 年 12 月)。MedDRA 中词的覆盖率远大于 WHO-ART,其适应面也更为广泛。MedDRA 中

低位语和首选语的关系除同义词关系外,有更大范围扩展,还包含了不同词法变体的关系、半同义词关系、亚元素关系等等。MedDRA 采用 8 位数字编码进行编码,每个层级的所有词都被赋予独一无二的 8 位数字编号,但数字本身无意义,从编码上看不出任何层级信息(参见下表 1.5),编码无解释性。

表 1.5 MedDRA 结构及层级关系
Table 1.5 The structure of MedDRA and relationship between different types of terms

Lowest Level	Preferred	High Level	High Level	System	Link
Term, LLT	Terms,PT	Terms,HLT	Group Terms, HLGT	Organ Class, SOC	
Diffuse vasculitis 10012978	Diffuse vasculitis 10012978	Vasculitides NEC 10047113	Vascular inflammations 10047116	Vascular disorders 10047065	Primary
Vasculitis diffuse 10047120		Vasculitides 10047112	Immune disorders NEC 10027665	Immune system disorders 10021428	Secondary

WHO-ART 和 MedDRA 虽有很多不同,二者也有相似之处。它们的首选语与上级的系统器官分类语(SOC)之间均存在多轴性,即在 SOC 层面上每个术语并不是完全互相排斥的。WHO-ART 和 MedDRA 中任何一个首选语都可能按不同的标准(如病理学、病因学、发病部位等)对应至不止一个 SOC。但任一个首选语只能通过一条 PT-HLT-HLGT 途径到达一个 SOC,且每个首选语都有一个对应的首选 SOC,以表 1.5 中词条为例,PT 术语 Diffuse vasculitis 的首选 SOC 为 Vascular disorders。二者亦均为多层次结构,为在不同层次上进行数据检索提供了便利。WHO-ART 与 MedDRA 之间已建立了基于文字匹配的映射关系,WHO-ART 中所有的 PT (Preferred Term) 都可以匹配到 MedDRA 中相应的条目。这种映射关系在每年 3 月份更新,此举方便了 2 个术语库之间的信息交流。

综合以上信息,WHO-ART 和 MedDRA 各具特色,WHO-ART 短小精悍,可从编码迅速判别词条的从属关系,且便于作为纸质版携带,对于小型机构而言

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库